DOCUMENT RESUME

ED 428 778 IR 057 326

AUTHOR	O'Daniel, Heather B.
TITLE	Cataloguing the Internet.
PUB DATE	1999-03-00
NOTE	6p.; "Associates" is a purely electronic journal (i.e., no printed equivalent).
AVAILABLE FROM	Web site: http://raven.cc.ukans.edu/~assoc/
PUB TYPE	Journal Articles (080)
JOURNAL CIT	Associates: the Electronic Library Support Journal; v5 n3 Mar 1999
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Authority Control (Information); Bibliographic Utilities; *Cataloging; Change; Classification; Cooperative Programs; *Indexing; Information Retrieval; *Internet; Library Automation; Library Catalogs; Problems; Quality Control; Subject Index Terms
IDENTIFIERS	Coalition for Networked Information; Dublin Core; *Electronic Resources; Library History; MARC; Metadata; OCLC

ABSTRACT

This paper discusses problems and opportunities, presented by the information explosion and the growth of the Internet, for libraries to apply and augment traditional methods of cataloging. The first section provides an overview of how the process of cataloging evolved, including the development of the Anglo-American Cataloging Rules (AACR), Library of Congress and Dewey Decimal classification systems, MARC format, OCLC, and Library of Congress Subject Headings. Issues or difficulties in applying classification systems to the information available on the Internet are explained in the second section, including lack of controlled vocabulary, lack of stability due to frequency of change to the data, and lack of quality standards. The third section shows the possibilities and plans for libraries to use cataloging for improving research on the Internet. Three current projects are described: (1) the Dublin Core, a set of metadata elements for cataloging electronic material; (2) the OCLC Cooperative Online Resource Cataloging Project, a research project exploring the cooperative creation and sharing of metadata in order to allow the integration of material available on the Internet with current library resources; and (3) the Coalition for Networked Information, a coalition of over 200 institutions and organizations that supports shared networked information resource and service development practices. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

IR

ED 428 778

Cataloguing the Internet

by

Heather B. O'Daniel
Intel Library,
Intel Corporation
heather.b.odaniel@intel.com

As it appeared in the
ASSOCIATES:The Electronic Library Support Staff Journal
Vol.5 No.3, March 1999

BEST COPY AVAILABLE

IR057326

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. B. O'Daniel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

The information explosion, with the creation of the Internet, has presented problems and opportunities for libraries to apply and augment traditional methods of cataloging. This research paper will cover three major topics, in an effort to explain some of the issues. The first topic will provide an overview of how the process of cataloging developed to establish an understanding of current systems. The second topic will explain issues or difficulties in applying classification systems to the information available on the Internet. And finally, the third topic will show the possibilities and plans for libraries to use cataloging for improving research on the Internet.

Since the advent of the printing press in the mid-15th century, mass-produced books have contained "conventions for representing information in published texts. Principle among these was the convention of the title page, which named the author and the title of the work contained therein, and also acknowledged the printing source (Tillett 2)." The key data of title, author, and source was then used to create the first bibliographic records.

Libraries began to place those bibliographic records into what was called a catalog. To catalog is to make a systemized list and so, the list of bibliographic records for the material housed in the library was called the catalog. Barbara B. Tillett explains that libraries first recorded lists in books. By the 1800's, the American Library Association had adopted the Anglo-American cataloging rules, published in a volume entitled AACR2, which is in use today. In 1901, the Library of Congress began selling printed cards to other libraries. Unlike book catalogs, card catalogs enabled the user to find the complete bibliographic description under many access points through the use of the newly-termed 'main entries' and 'added entries.' Main entries served as collating and arranging devices (Tillett 5). The ability to provide multiple access points developed the concept of indexing. Indexing used keywords or phrases to describe the content while pointing to the main entry or the bibliographic record.

In the 1800's, the Library of Congress classification system and the Dewey Decimal system were developed. Each system used letters and numbers to make up call numbers which represented the specific subject of a book. That allowed books to be organized on the shelf by subject matter ("Classification" 1). Because decimal numbers were used, the subject areas could easily be expanded using fractions of the whole numbers. In 1967, because of electronic databases, the Library of Congress converted bibliographic records into machine-readable cards or MARC. MARC format has five types of data: bibliographic, holdings, authority, classification, and community information. MARC records encode the data elements to help describe, retrieve, and control the information.

Another impact on the development of cataloging occurred in 1967, when a consortia called OCLC (Ohio College Library Center), formed a network of 54 Ohio Colleges using MARC records. In 1977, that network was opened to all libraries. In 1981, the legal name of the corporation became OCLC Online Computer Library Center, Inc. Today more than 30,000 libraries in the U.S. and other countries participate in the shared system ("History").

The ability to operate as a collective requires consistent standards for precise communication. An example is the word, movie. When referring to a book about the movie "Gone with the Wind", does a cataloger use moving picture, motion picture, cinema, film, or movie? To have consistent indexing requires an authority list or what may also be called a controlled vocabulary. The vocabulary list mentions each term, but states Motion Picture as the authority to be used in the record created.

The Library of Congress publishes a volume entitled the LC Subject Headings, which is accepted and used by most libraries. The volume lists the subject headings that are accepted for use when being cataloged. There are problems, though, when specialties require more precise categories. Some organizations publish a list of terms to provide the exact term used in a more concise subject classification. One such organization is Engineering Information, Incorporated, which has created a list called the Ei Thesaurus (Milstead).

So, this evolution has resulted in a system of collective consistency that each library classifies a book using the same key data, assigns keywords based on a controlled vocabulary, and places the records in a common database has enabled users to have quality results in the search for information.

With the advent of the Internet and the capability of sharing information electronically, the library world continues to evolve. The information explosion has increased the number of users, the amount of information available, and the speed of retrieval. This new direction causes problems in the attempt of library staff to apply traditional methods of cataloging. The search engines available on the Internet look for words in either the title, first few lines, or full text of the files. Searching can take too long and can produce results that have too many records, irrelevant records, or omissions to relevant records.

To perform cataloging of web sites requires consistent field entries similar to a MARC record. There are available fields within the programming language that make cataloging a viable idea. Within the Hypertext Markup Language (HTML) coding there is the ability to insert a field called a metatag. Metadata inserted into the metatag is similar to the information within a MARC record. Search engines may look specifically for matching terms in the metatag at amazing speed, but the terms input in the tags must be accurate. Today, web sites are thrown in the middle of the Internet without cataloging. It would be the same as just piling books in the center of a library with no system of indexing. The Internet lacks the structure of the library cataloging system.

This brings us to the first problem, which is controlled vocabulary. There is no source accepted by web creators that gives authority to the vocabulary words assigned to a site. Asking a web author to tag a site is like asking a book author to make his own MARC record after writing his book. This has always been the function of skilled librarians, using the common tools of authority lists, classification systems, or shared databases.

Other problems evolve when the information changes. If a book changes, it becomes a new edition with a new bibliographic record. Serials, also known as magazines, change frequently, but the change is predictable. In other words, the change could happen daily, monthly, or yearly, depending on the frequency of publication. The web sites on the Internet change erratically. Cataloging with a system using a main entry and

added entries would not work because there is no main entry. David Seaman, director of the Electronic Text Center at the University of Virginia/Charlottesville, pointed out, 'It's difficult to justify the time and expense of doing MARC cataloging of Internet materials on a large scale because what you have to catalog is so fluid. You go to the Web on a certain day and the item is there. Return in six months and it's not there. Or it's still there but has changed so dramatically that the record doesn't match anymore.' (Chepesiuk).

The final problem is quality standards. Authors approach a publisher who has a legal obligation and a professional reputation to produce a quality product. Librarians rely on consistent quality from reputable publishers to set the standards. One thing books had that resources on the Internet do not have is the accountability of a publisher. Publishers have a legal obligation to print the verifiable truth. They edit the content, structure, and grammar of their publications. They also verify the sources mentioned. So, this brings up the issue as to whether the Internet is even worth the time to catalog due to the varied quality.

There are three major problems in cataloging the Internet: the lack of universally accepted controlled vocabulary; the lack of stability due to frequency of change to the data; and the lack of quality standards.

There are many people trying to develop projects with the goal of establishing standards for all to use. The fact that there are so many efforts is a real problem in solidifying consistency. But there are three that seem to be getting the most attention, partly due to the institutions from which they started, the sponsorship, and the members.

Three main current projects include the Dublin Core, OCLC (CORC), and the Coalition for Networked Information (CNI).

In March 1995, fifty-two librarians, archivists, and scholars attended an OCLC-sponsored workshop to reach some agreement on what the core of a descriptive record for items on the Internet might include. The result was thirteen elements that they named the Dublin Core Metadata Element Set (Chepesiuk 60). The Dublin Core has become a prominent candidate for cataloging electronic material. Their goal was to create a set of metadata elements that, when defined, could be easily understood by web developers. Along with that basic ability, the elements provide the capability to further modify the data for more precise specialized communities of topics. The data elements selected include: title; author; subject; description; publisher; other contributor; date; resource type; format; identifier; source; language; relation; coverage; rights management.

Another OCLC effort is the Cooperative Online Resource Cataloging (CORC) Project. CORC is a research project exploring the cooperative creation and sharing of metadata by libraries. The goal is to allow libraries to integrate material available on the Internet with current library resources. According to Dorman, OCLC will build on the prior activities of NetFirst and InterCat, by seeding the initial CORC database with 145,000 records using full MARC and Dublin Core metadata (66).

Coalition for Networked Information (CNI) is another effort. "The goal of the coalition is to advance scholarship and intellectual productivity. Founded in 1990 by the Association of Research Libraries, Educom, and CAUSE. The members, who represent over two hundred institutions and organizations, meet bi-annually ("Coalition" 1). Bernbom informs that the coalition has created the Institution Wide Information Strategies project. Since each individual representative is gathering, delivering, and

storing electronic information, the strategic plan allows networked information resource and service development practices applicable to all (88).

Historically, the process of cataloging has proven a very effective method of organizing material for those seeking information. As the evolution of the electronic world continues, libraries have the opportunity to provide new ways of applying cataloging methods. As with all change, the transition can present problems, but the end result can be, hopefully, more than ever imagined.

Bibliography

Bernbom, Gerald. "Institution wide information strategies: a CNI initiative." Information Technology and Libraries June 1998:87-92.

Chepesiuk, Ron. "Organizing the Internet: The Core of the Challenge." American Libraries Jan. 1999:60-63.

"Classification Systems." Central Oregon Community College 1-2. Online. Internet. 12 Feb. 1999. Available http://www.cocc.edu/cfinney/lib127i/Organ.htm .

Coalition for Networked Information. Available http://www.cni.org .

CORC - Cooperative Online Resource Catalog. Available http://www.oclc.org/oclc/research/projects/corc/index.htm

Dorman, David. "Technically speaking: Can OCLC Do It Again?" American Libraries Dec. 1998:66.

"History of OCLC." OCLC Online Computer Library Center, Inc. n. pag. Online. Internet. 25 Jan. 1999. Available http://www.oclc.org/oclc/menu/history.htm .

Milstead, Jessica, ed. Ei Thesaurus 2nd ed. Hoboken: Engineering Information Inc., 1995.

Tillett, Barbara B. "Cataloging Rules and Conceptual Models." OCLC Distinguished Seminar Series 9 Jan. 1996:1-14. Online. Internet. 25 Jan. 1999. Available http://www.oclc.org:5046/~emiller/misc/tillett.html .